

○荒井隆行, 木下慶介, 程島奈緒, 楠本亜希子, 喜田村朋子 (上智大・理工)

1. はじめに

大きなホールで講演を聞く場合などでは音声聞きづらくなることがしばしばある。これは、話者の発する音声が聴取者に直接届く直接音ばかりではなく、室内におけるさまざまな反射の結果として重畳される残響の影響による。室内における音声明瞭度の客観的指標としては MTF (modulation transfer function) に基づく STI (speech transmission index) などが提案されている [1]。

残響環境下における音声明瞭度改善のための手法としては大きく2つのものに大別されるが、1つは残響が付加された音声に対して何らかの処理を施すことによって残響成分を抑圧する手法 (以下 post-processing) で、もう1つは残響が掛かる前に予め何らかの処理を施しておく手法 (以下 pre-processing) である。前者の post-processing については、単一のマイクロフォンを用いるものとして Langhans ら [2] によって提案された手法があり、さらに Avendano ら [3] によってその手法が拡張されている。いずれにおいても、音声を帯域に分割後、時間包絡に対するフィルタ処理を行い (以下、変調フィルタリング)、再び音声信号を再合成している。

一方、pre-processing に対してはやはり Langhans ら [2] によって同様の処理が提案されているが、ここでは顕著な明瞭度の改善が得られたとは報告されていない。その後、著者らはこの変調フィルタリングを用いて明瞭度の改善を試みた [4]–[7]。いずれの場合においても、変調フィルタは音声の syllabic rate を強調するように与えられ、結果として大幅な改善こそは得られなかったものの、何らかの改善傾向が確認されている。

本研究では、これらの変調フィルタリングによって得られる効果を詳しく調べるため、音声の定常部に対する抑圧処理に焦点をあてた。定常部に対して抑圧処理された音声信号は残響の影響を受けにくいことが予測されるが、それを実証するため、処理をする前と後の音声信号に対していくつかの残響を付加することによって、明瞭度への影響を調べると同時に、残響によるマスキングの影響について検討した。

2. 原理

残響が音声明瞭度を下げる説明としては、以下の仮説が立てられる。それは、残響は常に時間的に後方へ尾を引く結果、直前の音声信号に対する残響成分が現在の音声区間をマスクし、その現在注目する音声が聞きづらくなるというものである。これを表した例が図1の(a)と(b)である。図1の(a)の原信号に対し、(b)はその前半(1s前後まで)に対し残

響を畳み込んだ様子を示している。後半冒頭が特に前半からの残響の影響を強く受ける様子がわかる。

このように直前の音声区間に対し残響の尾が畳み込まれたものが現在の音声区間に重畳されるわけであるが、その影響を減らすためには直前の音声信号のエネルギーを抑えればよい。しかし、単にエネルギーを抑えるだけでは直前の音声区間自身の明瞭度が下がってしまう。そこで、明瞭度を損なわないようにエネルギーを減らす方法として、音声の遷移部はそのまま保存し定常部のみを抑圧する方法を考えた。

2.1 定常部抑圧処理

音声信号の定常部を抑圧するため、Furui (1986) にならって遷移の度合を表すパラメータ D を用いた [8]。Furui (1986) においては D はケプストラムの時間軌跡に対する回帰係数の2乗平均で定義されているが、本研究ではケプストラムは用いず、 $1/3$ -oct ごとに帯域制限された信号の時間包絡の対数に対する回帰係数の(帯域に渡る)2乗平均を用いた。なお、元の音声信号の標準化周波数は 16 kHz であったのに対し、時間包絡は 100 Hz にダウンサンプリングし、回帰係数は前後2点、計5点で計算している。

このようにして求められた D がある閾値を下回った場合を定常部とみなし、定常部と判定された区間の信号に対して元の振幅の40%にまで振幅を抑圧した。図1(c)は、(a)に対して定常部を抑圧した信号で、(d)は(c)の前半に対し残響を畳み込んだ結果である。図1の(b)と(d)を比較することによって、定常部抑圧処理を施した場合である(d)の方が、施さない(b)に比べて後方へ与えるマスキング量が少ないことがわかる。

3. 実験

定常部が抑圧された音声信号に対し残響を重畳した場合、処理による明瞭度低下はもたらされず後方へのマスキング量を減らすことが期待される。その効果を調べるために聴取実験を行った。

実験には明瞭度試験用の日本語単音節から20音節を用いた。ターゲットとなる単音節はキャリアセンテンス中に埋め込んだ。定常部抑圧処理としては、まったく行わないもの (no_proc)、ターゲットの直前まで処理を行ったもの (half_proc)、文全体を処理したもの (whole_proc) の3種類とした。残響に関しては弱いもの (残響時間=1.1s, 以下 Hall A) の強いもの (残響時間=1.8s, 以下 Hall B) 2種類を用いた。3種類の処理法に対して2種類の残響を畳み込んだもの、計6種類を刺激とした。被験者は12人で、1人は60刺激を聴取した。被験者はいずれも

* Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments.

By Takayuki Arai, Keisuke Kinoshita, Nao Hodoshima, Akiko Kusumoto and Tomoko Kitamura
(Sophia Univ., Tokyo, Japan)

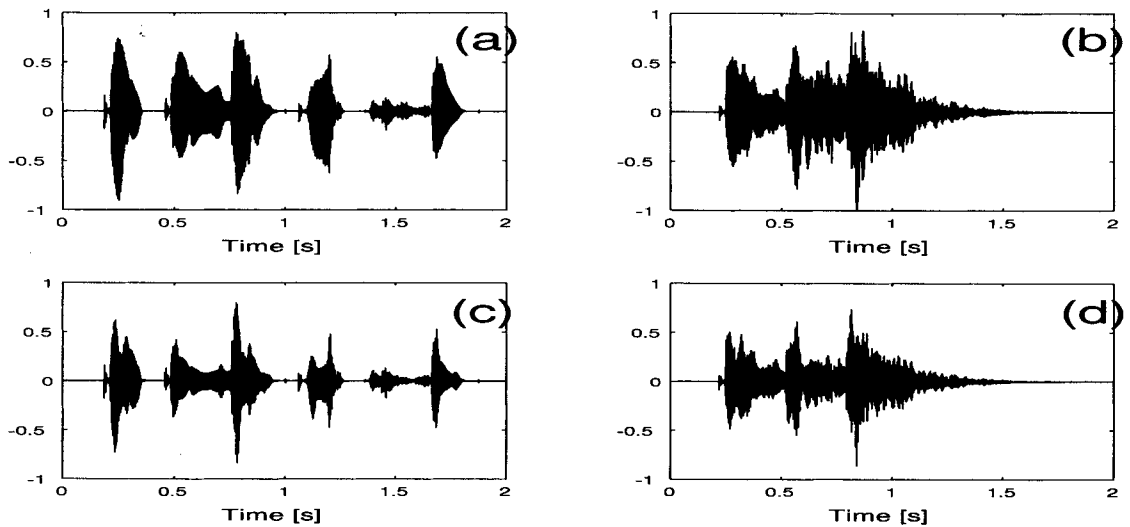


図 1. 時間波形の比較 (a) 原信号, (b) 原信号の前半に残響を畳み込んだもの, (c) 定常部抑圧信号, (d) 定常部抑圧信号の前半に残響を畳み込んだもの

健聴者であった。

表 1. % of correct responses.

	Hall A	B
no_proc	81.7	70.0
half_proc	87.5	72.5
whole_proc	85.8	60.8

4. 結果・考察

実験結果を表 1 に示す。この表を見るとわかるように、残響の程度が弱い Hall A に関しては、もとの音声信号に比べターゲットとなる単音節の直前まで処理を施した場合 (half_proc) の方が単音節明瞭度が高く、文全体を処理した場合 (whole_proc) でも、もとの音声信号よりも高い明瞭度を示した。あまり残響が強い Hall B の場合は、同様の傾向が示されなかった。

同様の処理を音韻バランス文 10 文に対して施し、同じように残響を付加することによって了解度を調べた。その結果、処理による了解度の低下は確認されなかった。

5. おわりに

定常部への抑圧処理をする前と後の音声信号に対していくつかの残響を付加することによって、明瞭度への影響を調べた。残響は常に時間的に後方へ尾を引く結果、直前の音声信号に対する残響成分が現在の音声区間をマスクし、その現在注目する音声聞きづらくなることが裏付けられた。変調フィルタリングは時間包絡の直流成分を抑圧していることから、間接的に各帯域における定常部抑圧を行っていることと見なせる。このことから、変調フィルタリングの効果も説明できる。

謝辞

残響のデータを提供して下さった東京大学の橋秀樹先生、上野佳奈子さん、横山栄さん、ならびに明瞭度試験用の音声を提供して下さった東京工業大学の古井貞熙先生にお礼申し上げます。また、聴取実験にご協力下さった被験者の皆様に感謝の意を表します。

参考文献

- [1] T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.*, Vol. 77, No. 3, pp. 1069-1077, 1985.
- [2] T. Langhans and H. W. Strube, "Speech enhancement by nonlinear multiband envelope filtering," *Proc. IEEE ICASSP*, pp. 156-159, 1982.
- [3] C. Avendano and H. Hermansky, "Study on the dereverberation of speech based on temporal envelope filtering," *Proc. ICSLP*, pp. 889-892, 1996.
- [4] 楠本亜希子, 荒井隆行, 喜田村朋子, 高橋真保呂, 村原雄二, "室内音響における聴覚障害者のための音声処理に関する検討," 日本音響学会講演論文集, Vol. 1, pp. 389-390, Sep. 1999.
- [5] A. Kusumoto, T. Arai, T. Kitamura, M. Takahashi, and Y. Murahara, "Modulation enhancement of speech as a preprocessing for reverberant chambers with the hearing-impaired," *Proc. IEEE ICASSP*, Vol. 2, pp. 853-856, 2000.
- [6] 喜田村朋子, 荒井隆行, 楠本亜希子, 村原雄二, "聴覚障害者のための残響にロバストな音声処理—変調フィルタの検討—," 日本音響学会講演論文集, Vol. 1, pp. 333-334, Mar. 2000.
- [7] T. Kitamura, K. Kinoshita, T. Arai, A. Kusumoto and Y. Murahara, "Designing modulation filters for improving speech intelligibility in reverberant environments," *Proc. ICSLP*, Vol. 3, pp. 586-589, 2000.
- [8] S. Furui, "On the role of spectral transition for speech perception," *J. Acoust. Soc. Am.*, Vol. 80, No. 4, pp. 1016-1025, 1986.